Journal of Surgical Oncology 64:135–142 (1997)

Measuring the Ability of Residents to Manage Oncologic Problems

DAVID A. SLOAN, MD,* MICHAEL B. DONNELLY, PhD, RICHARD W. SCHWARTZ, MD, PATRICK C. McGRATH, MD, DANIEL E. KENADY, MD, DAVID P. WOOD, MD, AND WILLIAM E. STRODEL, MD
Department of Surgery, University of Kentucky Chandler Medical Center, Lexington, Kentucky

Background: An OSCE was used to measure the ability of a cohort of residents to manage oncologic problems.
Methods: Nine oncologic clinical problems were presented to 56 surgical residents. Each problem contained a 5-minute data-gathering period (DGP) and a 5-minute data-interpretation period (DIP). A performance score was determined for each resident for each problem. Reliability was estimated by coefficient α; validity, by the construct of experience. Wilks's λ criterion was used to determine whether training level could be identified by OSCE performance.
Results: The DGP reliability was 0.80; the DIP, 0.49. Senior residents performed significantly better than junior residents ($P = 0.0001$), who performed significantly better than interns ($P = 0.0009$). Of the residents, 62% were competent on the DGP, but only 21% on the DIP. Important deficits in knowledge and clinical skills were apparent at all levels of training.
Conclusion: The education and evaluation of residents in oncology need improvement. J. Surg. Oncol. 64:135–142.

KEY WORDS: clinical competence; objective structured clinical examination; cancer education; graduate medical education

INTRODUCTION

Faculty members rarely observe residents interacting with cancer patients. Furthermore, evidence at both the undergraduate and the graduate medical education levels suggests that students and residents receive little education, formal or informal, in oncology [1–7]. As a result of recent changes in the structure of the health care system, many cancer patients are never seen by surgical residents until the day of surgery, by which time important treatment decisions already have been made independently by faculty members. On the basis of pilot data from our institution showing important deficits in the ability of surgical interns to solve oncology problems [5], we elected to use performance-based testing to measure the ability of an entire cohort of residents to manage a more extensive series of such problems. A striking advantage of the Objective Structured Clinical Examination (OSCE) [8–11] format used in this study of residents' performance is that it places individual residents face-to-face with actual patients, allowing faculty members an excellent opportunity to examine the residents' clinical, interpersonal, and problem-solving abilities.

The purpose of this study was to determine the reliability, construct validity, and educational utility of a

This report was presented at the annual meeting of the Society for Surgical Oncology, Houston, Texas, March 17–20, 1994.

*Correspondence to: Division of General Surgery, Section of Surgical Oncology, University of Kentucky Chandler Medical Center, 800 Rose Street, Lexington, KY 40536-0084.

Accepted 19 October 1996

nine-problem OSCE designed to assess the oncologic clinical skills of general surgery residents.

MATERIALS AND METHODS

Nine oncologic clinical problems were presented in an OSCE format to 56 surgical residents at the University of Kentucky as part of a larger 19-problem examination. Postgraduate year (PGY) levels ranged from PGY-1 through PGY-6. Each of the surgical residents was presented with the same nine cancer patients. Four of these were actual cancer patients; five were simulated patients. The actual patients included a patient with fibrocystic disease and a mammogram showing suspicious microcalcifications; a patient with a pigmented skin lesion (to rule out melanoma); a patient with a malignant thyroid nodule; and a patient with hypercalcemia. Simulated problems included a patient with a history suggestive of colon cancer; a patient seeking an opinion on treatment options after a breast biopsy positive for cancer; a patient with a new onset of cough and a chest radiograph suggestive of malignancy; a patient with a history highly suggestive of tongue cancer; and a patient with bladder cancer who requested treatment for hematuria. Each patient, whether actual or simulated, was trained by a standardized patient trainer to ensure standardization of response to both questioning and physical examination by trainees. One hour of training per patient was generally sufficient.

The interaction of the resident with each of these individual patients was highly structured. Each resident was allowed 5 minutes for the patient interaction. This initial 5-minute interaction was called the data-gathering period (DGP). During this time, a faculty member observed the residents and compared their performance against a checklist of predetermined items that had been deemed essential by faculty experts. For each of the items deemed important in this initial clinical evaluation, three levels of performance were possible: (1) not done, (2) done poorly, and (3) done well. Figure 1 shows a checklist for the tongue cancer patient. Figure 2 shows sample evaluation items for the various oncologic problems. Faculty experts were assigned the task of preparing the checklists; generally, input was obtained from additional faculty members with similar expertise. Faculty members were briefed both in person and in writing before the examination. Most of the faculty members had had prior experience with the OSCE format. For each of the nine problems, one or at most two faculty members examined the entire cohort of trainees.

Immediately upon completion of the DGP, each resident was presented with a series of several short-answer questions that dealt directly with the information gleaned from the patient during the DGP. Frequently, a radiograph was also viewed at this time. The residents were allowed 5 minutes to respond to the questions, and their written responses were later graded by faculty experts, again according to preset objective criteria. This second 5-minute period was labeled the Data Interpretation Period (DIP). A performance score (the percentage of the indicated history or physical examination items obtained or the completeness of the diagnosis and treatment plan) was determined for each resident for each clinical problem.

Nineteen faculty members were involved with the grading. In general, two faculty members were involved with each individual patient problem. The faculty members generally divided their responsibilities between the 2 days on which the examination was conducted. The faculty members involved in proctoring the examination and in developing the individual patient problems were not exclusively from the Department of Surgery; two were from the Department of Internal Medicine. The surgery faculty members represented a number of subspecialties, including surgical oncology, gastrointestinal surgery, head and neck surgery, plastic surgery, urologic oncology, and thoracic surgery.

The reliability of the examination was estimated by coefficient α. Reliability is an indicator of the precision of a measurement instrument (in this case, the OSCE) and varies from 0.00 to 1.00. A reliability of .80 is considered desirable. The precision of the measurement instrument decreases as the reliability decreases. A reliability of .70 is considered adequate. Reliabilities of <.70 reflect decreasing precision. Validity was assessed by using the construct of experience, i.e., if the examination is valid, senior residents should perform better than junior residents. For the purposes of data analysis, the residents were grouped as follows: (1) incoming interns ($n = 18$); (2) junior residents (PGY-2 or -3 residents, $n = 25$); and (3) senior residents (PGY-4, -5, or -6 residents, $n = 13$). A stepwise multiple-group discriminant function, using Wilks's λ criterion, was used to determine the degree to which the clinical performance of the three groups of residents could be differentiated by their performance on both the DGP and the DIP sections of the OSCE. Multiple-group discriminant function identifies an optimal combination of scores that significantly differentiate the groups. Its overall statistical significance is tested by the Chi-square statistic.

RESULTS

The reliability of the nine DGPs was 0.80; that of the DIPs, 0.49. One of the two possible discriminant functions comparing the performance of the three resident groups on the nine DGP measures was significant ($\chi^2 = 55.06$; $df = 10$; $P < .0001$). Table I summarizes the results of this analysis. The group centroid (the average standardized score across problems listed at the bottom of Table I) and the means for each problem demonstrate that the interns performed worst, and the senior residents,

UNIVERSITY OF KENTUCKY DEPARTMENT OF SURGERY
OSCE EXAMINATION

(revised 6/93) Head and Neck Station
Resident 93/94 HN-A (Place name tag here)

Part A

Mr. Osborn comes to you complaining of a sore on his tongue. Please take a focused history.

	Not done	Done Poorly	Done Well
Asks age	0	1	2
Location of sore on his tongue	0	1	2
Type of discomfort	0	1	2
Duration of discomfort	0	1	2
Trauma history	0	1	2
Difficulty with eating/swallowing	0	1	2
Sore throat/pain on swallowing	0	1	2
Spitting blood	0	1	2
Tongue numbness	0	1	2
Ear pain	0	1	2
Hoarseness	0	1	2
Anorexia	0	1	2
Weight loss	0	1	2
Cough	0	1	2
Dental history	0	1	2
Awareness of associated neck masses	0	1	2
Alcohol history	0	1	2
Tobacco use			
cigarettes (amount/per year)	0	1	2
oral tobacco	0	1	2
snuff	0	1	2
Past medical history (COPD, cardiac)	0	1	2
History of TB, syphilis	0	1	2
History of head and neck cancer	0	1	2
History of other cancer (including skin)	0	1	2
Medications	0	1	2

Overall Evaluation:	**Not at All**				**Very Much**
Organized approach to performing the physical history	0	1	2	3	4
Interacted effectively with the patient	0	1	2	3	4
	Poor		**Average**		**Outstanding**
Overall Evaluation	0	1	2	3	4
Competent		**NO**		**YES**	
Overall by the patient:	**Poor**				**Outstanding**
Interpersonal evaluation	0	1	2	3	4

Fig. 1. Data-gathering checklist for the evaluation of a patient with tongue cancer.

SAMPLE CHECKLIST ITEMS

Thyroid Cancer (of 25 items)
- Asks about change in voice
- Asks about difficulty in swallowing
- Asks about neck irradiation

Bladder Cancer (of 22 items)
- Asks about blood in urine (color, duration, frequency)
- Asks about pain with urination
- Asks about blood-thinning medications

Breast Pain/Microcalcifications (of 20 items)
- Elicits nature of pain
- Asks about nipple discharge
- Elicits family history

Colon Cancer (of 22 items)
- Asks about nature of onset of pain
- Asks about palliating/aggravating factors
- Asks about anorexia/weight loss

Breast Cancer Options (of 15 items)
- Places patient at ease
- Explains that there is more than one option
- Emphasizes that breast preservation is a very good option with results equivalent to those achieved by mastectomy

Melanoma (of 16 items)
- Asks about change in lesion
- Asks about itching of lesion
- Asks about bleeding

Lung Cancer (of 22 items)
- Asks about frequency of cough
- Asks about presence of sputum
- Notes raspy quality of voice

Hypercalcemia (of 22 items)
- Asks about fatigue
- Asks about constipation
- Asks about weight change

Fig. 2. Sample checklist items for the oncologic problems.

TABLE I. Discriminant Function Analysis Comparing Levels of Training for DGP† Measures

Problem	Mean intern score	Mean jr. resident score	Mean sr. resident score	Univariate F	Step in function	Correlation with function
Colon cancer	70	71	80	5.75*	4	.33
Breast cancer options	43	67	84	26.68*	1	.77
Breast microcalcifications	64	69	78	7.41*	3	.41
Lung cancer	79	81	86	1.88		.23
Tongue cancer	46	54	68	14.10*	5	.57
Melanoma	59	65	76	12.32*	2	.53
Thyroid cancer	46	50	62	7.39*		.49
Bladder cancer	51	53	55	0.39		.31
Hypercalcemia	49	51	63	6.03*		.36
Group centroid	−1.43	0.02	1.94			

†DGP = Data-Gathering Period; the 5-minute period during which residents gathered history information directly from actual or simulated oncology patients.
*$P < 0.01$.

best. Pair-wise F-ratios comparing the performance of these three groups indicated that the senior residents performed significantly better than the junior residents ($F = 6.62$; df = 5,49; $P = .0001$) and the interns ($F = 15.87$; df = 5,49; $P < .0001$); in turn, the junior residents performed significantly better than the interns ($F = 4.96$; df = 5, 49; $P = .0009$).

The univariate F-ratios (fifth column of Table I) for each problem indicated significant differences in performance levels among the three groups on seven of the nine problems. The breast cancer options scenario was the most differentiating, whereas the lung cancer and bladder cancer problems did not significantly differentiate the groups. Performance on the lung cancer problem was good; in contrast, performance on the bladder cancer station was poor at all levels of training. A composite score, consisting of scores on five of the problems (column six of Table I), led to an optimal differentiation of the three groups. Column 7 of Table I presents the correlation of each of the DGP scores with the discriminant function (the higher the correlation, the more a problem differentiated the groups). The latter two analyses indicate that the breast cancer options problem and to a lesser degree, the tongue cancer problem best differentiated the groups, whereas the lung cancer and bladder cancer problems least differentiated them.

The profiles of the five DGP scores included in the discriminant function analysis allowed a correct group classification in 71% of the cases (Table II). Table II also indicates that each of the senior residents and interns had a distinctive score profile. In contrast, a number of the junior residents performed as well as the senior residents, and a number performed as poorly as the interns.

The summary of the discriminant function analysis of the DIP scores is presented in Table III. Both of the two discriminant functions were significant (Function 1: $\chi^2 = 62.55$; df = 14; $P < .0001$; Function 2: $\chi^2 = 14.73$; df = 6; $P = .0224$). Six of the OSCE problems added significantly to the differentiation of the three groups. The group centroids (listed for each group at the bottom of Table III) indicate that the first function differentiated the senior residents from the junior residents and the interns, whereas the second function separated the interns from the junior and senior residents. The breast cancer options problem and the thyroid cancer problem best differentiated the senior residents from the junior residents and interns (column 7 of Table III), whereas the breast microcalcifications problem and the tongue cancer problem best differentiated the interns from the junior and senior residents (column 8 of Table III). The univariate F-ratios (column 5 of Table III) indicated that six of the nine stations significantly differentiated the three groups. There were no significant differences among the groups on the colon cancer, melanoma, and bladder cancer stations; performance was poor on all three of these problems. As with the DGP, the senior residents performed better than the junior residents ($F = 10.49$; df = 7, 47; $P < .0001$) and interns ($F = 7.03$; df = 7, 47; $P < .0001$). The junior residents performed better than the interns ($F = 3.23$; df = 7, 47; $P = .0071$).

On the basis of the six DIP measures included in the

TABLE II. Actual and Predicted Group Membership on the Basis of DGP* Measures

Group	Predicted group		
	Intern	Junior resident	Senior resident
Intern	78%	22%	0%
Junior resident	24%	56%	20%
Senior resident	0%	8%	92%
(71% correctly classified)			

*DGP = Data-Gathering Period; the 5-minute period during which residents gathered history information directly from actual or simulated oncology patients.

TABLE III. Discriminant Function Analysis Comparing Levels of Training for DIP† Measures

Problem	Mean intern score	Mean jr. resident score	Mean sr. resident score	Univariate F	Step in function	Correlation with function 1	Correlation with function 2
Colon cancer	51	52	54	0.41	-	.01	.19
Breast cancer options	32	33	50	6.94*	4	.44	.02
Breast microcalcifications	35	53	58	8.35*	1	−.02	.68
Lung cancer	54	65	77	6.32*	7	.22	.34
Tongue cancer	42	52	56	7.04*	2	.06	.59
Melanoma	66	60	71	2.10	5	.32	−.26
Thyroid cancer	53	53	71	5.78*	6	.43	−.02
Bladder cancer	61	63	65	0.19	-	.15	−.23
Hypercalcemia	46	49	66	5.26*	3	.35	.10
Group centroid 1	−0.66	−0.54	1.96				
Group centroid 2	−1.16	0.39	0.86				

†DIP = Data Interpretation Period; the 5-minute period during which residents wrote answers to questions about their interaction with oncology patients.
*$P<0.01$.

discriminant function, the members of the three groups could be identified with 77% accuracy. Table IV presents the observed and predicted group membership. As can be seen, 83% of the interns were correctly classified; none had a profile similar to that of the senior residents. Of the senior residents, 92% had the profile of a senior resident, and none had the profile of the intern group. In contrast, the profile of the junior residents was again less clearly defined. Only 64% of this group had similar profiles: 24% had profiles similar to the interns, and 12% had profiles similar to the senior residents.

The evidence shows that the reliability and construct validity of this examination are good, but what does this examination tell us about the competence of the residents in relation to surgical oncology? If we operationally define clinical competence as a score of 60% or higher, then, 62% of the residents were competent on the DGP, but only 21% were competent on the DIP. More encouraging is the fact that residents steadily improve their clinical skills as they move through the surgical training program.

TABLE IV. Actual and Predicted Group Membership on the Basis of DIP* Measures

Group	Predicted group		
	Intern	Junior resident	Senior resident
Intern	83%	27%	0%
Junior resident	24%	64%	12%
Senior resident	0%	8%	92%
(77% correctly classified)			

*DIP = Data Interpretation Period; the 5-minute period during which residents wrote answers to questions about their interaction with oncology patients.

DISCUSSION

The treatment of cancer patients is becoming more complex as more and more solid tumors are treated in a multidisciplinary fashion, requiring the input of multiple cancer specialists. As an example, the treatment of a breast cancer patient may involve an oncologic surgeon, a medical oncologist, a radiotherapist, a radiologist, a pathologist, and a nurse oncologist. Often it is no longer possible or, for that matter, appropriate for individual surgeons alone to treat cancer patients, even when treatment begins while the cancer is in a very early stage. Furthermore, rarely does a patient with a solid tumor have only a single option; generally, multiple options and multiple treatment timetables are possible.

The multitude of options and the burgeoning body of oncologic knowledge make medical and surgical education in cancer management difficult at both the undergraduate and the graduate levels. Many medical schools lack an integrated oncology curriculum [12,13]; reviews of undergraduate medical curricula have documented many weaknesses in the presentation of oncology [2,12]. An overview of undergraduate education in 18 medical schools showed that there was a distinct lack of educational activity in general oncology [12]. This lack of a structured oncology curriculum may explain the poor performance of medical students in oncology-related areas. Bleyer et al. [2] noted that medical students' performance on oncology-related questions was significantly lower than their performance on questions from other disciplines. In examining oncology education at the University of Washington School of Medicine, these authors noted that oncology education was, in general, poor. It was particularly ineffective in oncology topics most relevant to general surgery [2]. Jewell et al. [1] evaluated the presentation of surgical oncology at the University of Kansas Medical Center. Their review of the curriculum revealed that students could graduate from medical school without any appreciable insight into the care of

the cancer patient [1]. A recent survey of fourth-year medical students at three institutions revealed that most students had never palpated a cancerous breast [4]. To cope with these and other educational needs, Mirand [14] has recommended that every comprehensive cancer center have a full-time educational director. He emphasizes the central importance of education in the mission of the cancer center [14].

Although it is the subjective impression of many authors that both medical students and residents have numerous deficits in their appreciation of oncology problems, very few published data exist to support this contention. There is some subjective comment that the area of poorest performance for surgeons taking the board examination is surgical oncology problem solving [3]. A previous study at our institution found that students and interns demonstrated serious deficits in oncologic problem solving when they were presented with several common clinical problems [5]. Mean performance scores in that study were very poor, ranging from 32% to 72%. When 70% was set as a passing score in that study, most students and interns failed almost all of the problems. The deficits identified encompassed the areas of information gathering, diagnosis, and treatment [5]. To the best of our knowledge, that pilot study was the first to use performance-based testing to evaluate the ability of interns and students to treat cancer patients.

The advantage of the OSCE format [8–11], which relies heavily on the use of patients, is that it provides an extra dimension in the assessment of residents [9–11]. We recently demonstrated that the OSCE provides information about residents' clinical abilities that is not available when customary evaluation tools, such as the ward evaluation, the in-training examination, and the oral examination, are used [15,16]. The OSCE allows us to present the same clinical problem to a multitude of individuals in a highly structured setting in which faculty can directly evaluate the residents according to *objective* criteria. The OSCE thus provides a unique window into the clinical ability of individual residents. For example, the opportunity for a faculty member to witness residents taking a history from and examining an actual patient with a malignant thyroid mass is most valuable. Although the residents' knowledge base is important, the faculty member also is able to witness their interpersonal skills, their ability to interview efficiently, and their ability to identify pertinent positive and negative physical findings. No amount of multiple-choice questioning can duplicate this scenario. Another advantage of the OSCE format is that it forces faculty members to determine which clinical items related to an appropriate history and physical examination are *essential* to a *competent* performance.

The OSCE sends a clear message to residents that faculty members strongly emphasize the interviewing and examination of cancer patients. Abilities other than history-taking and physical examination can be evaluated as well, as in the scenario of the patient seeking advice about whether to have a mastectomy. This study showed that, although there were indeed differences in levels of performance between junior and senior residents, there were still very important deficits in oncologic problem solving at the more senior levels. The authors are convinced that the use of actual and simulated patients in a structured setting provides an unparalleled opportunity to witness the interaction of residents with cancer patients. This innovative form of testing needs to be expanded so that faculty members can gain a greater appreciation of where they are failing in their education of residents.

Very few standards have been set with regard to the evaluation of cancer patients. Conversely, for patients with acute trauma or acute cardiac disease, appropriate groups of specialists have laid out and encoded descriptions of competent behavior (e.g., the Advanced Cardiac Life Support and Advanced Trauma Life Support [ATLS] programs) [17,18]. It is indeed notable that a successful, large-scale program like ATLS stems from the subjective impression of a single concerned physician that there was widespread incompetence in the initial treatment of the injured patient [18]. Such a comprehensive, *performance-based* educational program does not exist for the treatment of cancer patients, who arguably present a much more complicated set of management problems. The authors believe that the time for such an educational intervention has arrived. In fact, an NIH-funded, multi-institutional study based at our institution is underway to determine the effectiveness of a modification of the OSCE for instructing residents in the multidisciplinary skills involved in the treatment of breast cancer patients.

CONCLUSIONS

In conclusion, the use of actual and simulated patients in structured clinical settings provides an excellent opportunity for witnessing the interaction of residents with cancer patients. As expected, more experienced residents perform at a higher level than more junior residents. Nonetheless, important deficits in patient interaction and problem solving are clearly seen even at the highest levels of training. Our contention is that both surgical and medical oncologists need to assume greater responsibility for the education and evaluation of residents in the disciplines of oncology.

REFERENCES

1. Jewell WR, Deslauriers MP: Teaching surgical oncology to medical students. J Surg Oncol 1983;24:218–220.
2. Bleyer WA, Hunt DD, Carline JD, et al.: Improvement of oncology education at the University of Washington School of Medicine, 1984–1988. Acad Med 1990;65:114–119.

3. Balch CM: Surgical oncology in the 21st century. Arch Surg 1992;127:1272–1277.
4. Kennedy BJ, Foley JF, Yarbro JW, Ringenberg QS: Breast cancer examination experience of medical students and residents. J Cancer Ed 1988;3:243–245.
5. Sloan DA, Donnelly MB, Schwartz RW, et al.: Assessing medical students' and surgery residents' clinical competence in problem solving in surgical oncology. Ann Surg Oncol 1994;1:204–212.
6. Bernard LJ, Carter RA: Surgical education in U.S. and Canadian medical schools. J Cancer Educ 1988;3:239–242.
7. Smith WT, Tattersall MH, Irwig LM, Langlands AO: Undergraduate education about cancer. Eur J Cancer 1991;27:1448–1453.
8. Harden RM, Stevenson M, Downie WW, Wilson GM: Assessment of clinical competence using objective structured examination. Br Med J 1975;1:447–451.
9. Sloan DA, Donnelly MB, Johnson SB, et al.: Use of an Objective Structured Clinical Examination (OSCE) to measure improvement in clinical competence during the surgical internship. Surgery 1993;114:343–351.
10. Cohen R, Reznick RK, Taylor BR, et al.: Reliability and validity of the objective structured clinical examination in assessing surgical residents. Am J Surg 1990;160:302–305.
11. Petrusa ER, Blackwell TA, Ainsworth MA: Reliability and validity of an objective structured clinical examination for assessing the clinical performance of residents. Arch Intern Med 1990;150:573–577.
12. Brincker H: Undergraduate oncology education in Scandinavia. J Cancer Ed 1988;3:97–101.
13. Robert KH, Einhorn J, Kornhuber B, et al.: European undergraduate education in oncology. A report of the EORTC Education Branch. Acta Oncol 1988;27:423–425.
14. Mirand EA: Education in a comprehensive cancer center. Cancer 1972;4:863–865.
15. Schwartz RW, Donnelly M, Sloan D, et al.: Assessing senior residents' knowledge and performance: an integrated evaluation program. Surgery 1994;116:634–640.
16. Schwartz R, Donnelly M, Sloan D, et al.: The relationship between faculty ward evaluations, OSCE, and ABSITE as measures of surgical intern performance. Am J Surg 1995;169:414–417.
17. Carveth SW, Burnap TK, Bechtel J, et al.: Training in advanced cardiac life support. JAMA 1976;235:2311–2315.
18. Collicott PE: Advanced Trauma Life Support (ATLS): past, present, future—16th Stone Lecture, American Trauma Society. J Trauma 1992;33:749–753.